



SMW Multipage Spreadsheet Editing

SMWCon Fall 2021

Tim Holzheim / RWTH Aachen i5
Wolfgang Fahl / RWTH Aachen i5

Semantic MediaWiki Multipage (SMW) Spreadsheet Editing

- Why do most SMW sites split 1:n relations as master/slave pages?
- Why is the master/slave pages approach slow for editing?
- The double-edged sword of CSV import
- Improved approach: spreadsheet editing
- OpenResearch use case
- What about a generic spreadsheet editing solution?
- Discussion

- Main page data is typically presented in InfoBox
 - Data is mostly in Mediawiki Macro Markup - WikiSON (Wiki Simple Object Notation)
- Our use case:
 - Entities where each entity has its own page
 - 1:n relations between entities
 - E.g Series has n Events
 - Entities properties are defined via the WikiSON parameters

Event:

SMWCon 2020	
Semantic MediaWiki Conference	
Event in series	SMWCon
Dates	2020/11/24 (iCal) - 2020/11/26
Homepage:	https://www.semantic-mediawiki.org/wiki/SMWCon Fall 2020
Twitter account:	@SemanticMW

WikiSON markup:

```
{{Event
|Acronym=SMWCon 2020
|Title=Semantic MediaWiki Conference
|Series=SMWCon
|Ordinal=17
|Type=Conference
|Start date=2020/11/24
|End date=2020/11/26
|Submission deadline=2020/10/15
|Twitter account=@SemanticMW
|presence=online
}}
```

Why do most SMW sites split 1:n relations as master/slave pages?

Alternatives are page per child or subobject per child ...

- [List of Problems](#)
- Subobject handling in SMW is not loved much
 - [Complicated queries](#)
- PageForms will give some useability but not comparable to what a spreadsheet can do
- No proper spreadsheet mode available (as far as we know ...)
- See e.g. <https://stackoverflow.com/questions/1680747/using-semantic-mediawiki-for-tabular-data>

List of Problems

Icon	Topic	name	pluralName	documentation
	Problem	Problem	Problems	I am a documentation for a problem/solution cycle.

Problems [\[edit\]](#)

39 Problems:

Problem	Problem	at	project	state	Diagnosis	Therapy	followup	done	effort	workpackage
Workdocumentation 2021-03-24 TH/Problems	Migrating City, State, Country pages to City, Region, Country pages in which the data in encoded as a template See • Concept:City • Concept:Region • Concept:Country	25 March 2021 09:46:00	ConfIDent	InProgress	Current Cities, states and countries just had the template ""UNIQ--nowiki-00000005-QINU"" ... Information about the name of the corresponding location where taken from the page name.	1. wikibackup off all location entities 2. wikibackup -s or -q ""UNIQ--nowiki-00000006-QINU"" --backupPath or migrate_city 3. wikibackup -s or -q ""UNIQ--nowiki-00000007-QINU"" --backupPath or migrate_region 4. wikibackup -s or -q ""UNIQ--nowiki-00000008-QINU"" --backupPath or migrate_country 5. Running a script that updates all files by replacing the current template with ""UNIQ--nowiki-00000009-QINU"" ... 6. wikirestore to upload the migrated file to or migrate		24 March 2021		ConfIDent WP2
Workdocumentation 2021-03-24 TH/Problems	Adding a command line interface to the implemented templates to be able to generate wikispaces from a given input.	26 March 2021 07:27:00	ConfIDent	Done						
Workdocumentation 2021-03-25 TH/Problems	Test the generator in wikirender for: • Concept:Project • Concept:Workpackage • Concept:Task	26 March 2021 07:33:00	ConfIDent	Done		1. Get the needed data from this wiki 1. Topics: \$ wikiquery -s rq -q ""UNIQ--nowiki-00000002-QINU""?mainlabel=Topic! ?Topic name = name! ?Topic pluralName = pluralName! ?Topic icon = icon! ?Topic iconUrl = iconUrl! ?Topic documentation = documentation! ?Topic wikiDocumentation = wikiDocumentation! ?Topic defaultstoremode = defaultstoremode! ?Topic listLimit = listLimit! ?Topic cargo = cargo! ?Topic headerTabs = headerTabs! ?Topic context = context! limit=200" -> rq_topics.json 2. Remove the unwanted topics from the result file 2. Properties: \$ wikiquery -s rq -q ""UNIQ--nowiki-00000003-QINU""?mainlabel=Property! ?Property name = name! ?Property label = label! ?Property type = type! ?Property index = index! ?Property sortPos = sortPos! ?Property primaryKey = primaryKey! ?Property mandatory = mandatory! ?Property namespace = namespace!				

Edit Problem: Workdocumentation 2021-03-24 TH/Problems

Problem [\[edit\]](#)

...	Migrating City, State, Country pages to City, Region, Country pages in ... · 2021-03-25T09:46:00.000Z · 25 · Mar · 2021 · 09 · 46 · 00 · ConfIDent · ConfIDent WP2 · Ir	...
...	Adding a command line interface to the implemented templates to be abl... · 2021-03-26T07:27:00.000Z · 26 · Mar · 2021 · 07 · 27 · 00 · ConfIDent · Done	...
<input type="button" value="Add another"/>		

Why is the master/slave approach slow for editing?

- Each page requires a manual edit
 - Min. 4 clicks/page
 - For this Example ~110 clicks to add the ordinal

Ordinal	Year	From	To	City	Country	presence	Homepage	TibKatid	GND	dblp	WikiCFP	Wikidata
WWW 1995	Error: Invalid time.						http://iw3c2.org					
WWW 2020	2020	Apr 20	Apr 24	TW/TPE/Taipei	TW		https://www2020.thewebconf.org/					
WWW 2019	2019	May 13	May 17	US/CA/San Francisco	US		https://www2019.thewebconf.org/					
WWW 2018	2018	Apr 23	Apr 27	FR/ARA/Lyon	FR		https://www2018.thewebconf.org/					
WWW 2017	2017	Apr 3	Apr 7	AU/WA/Perth	AU		http://www.www2017.com.au					
WWW 2015	2016	May 18	May 22	IT/52/Florence	IT							
WWW 2016	25	Apr 11	Apr 15	CA/QC/Montreal	CA		http://www2016.ca/					
WWW 2014	2014	Apr 7	Apr 11	KR/11/Seoul	KR		http://www2014.kr					
WWW 2013	2013	May 13	May 17	BR/RJ/Rio de Janeiro	BR		http://www2013.org					
WWW 2012	2012	Apr 16	Apr 20	FR/ARA/Lyon	FR		http://wwwconference.org/www2012					
WWW 2011	2011	Mar 28	Apr 1	IN/TG/Hyderabad	IN		http://wwwconference.org/www2011					
WWW 2010	2010	Apr 26	Apr 30	US/NC/Raleigh	US		http://www2010.org				6181	
WWW 2009	2009	Apr 20	Apr 24	ES/MD/Madrid	ES		http://www2009.org					
WWW 2008	2008	Apr 21	Apr 25	CN/BJ/Beijing	CN		http://www2008.wwwconference.org					
WWW 2007	2007	May 8	May 12	CA/AB/Banff	CA		http://www2007.wwwconference.org					
WWW 2006	2006	May 22	May 26	GB/SCT/Edinburgh	GB		http://iw3c2.org					
WWW 2005	2005	May 10	May 14	JP/12/Chiba	JP		http://www2005.wwwconference.org					
WWW 2004	2004	May 17	May 22	US/NY/New York City	US		http://www2004.wwwconference.org					
WWW 2003	2003	May 17	May 22	HU/PE/Pest	HU		http://www2003.wwwconference.org					
WWW 2002	2002	May 6	May 11	US/HI/Honolulu	US		http://www2002.wwwconference.org					
WWW 2001	2001	May 1	May 5	CN/HK/Hong Kong	CN		http://www10.wwwconference.org					
WWW 2000	2000	May 15	May 19	NL/NH/Amsterdam	NL		http://www9.wwwconference.org					
WWW 1999	1999	May 11	May 14	CA/ON/Toronto	CA		http://www8.wwwconference.org					
WWW 1998	1998	Apr 14	Apr 18	AU/QLD/Brisbane	AU		http://www7.wwwconference.org					
WWW 1997	1997	Apr 7	Apr 11	US/CA/Santa Clara	US		http://iw3c2.org					
WWW 1996	1996	May 6	May 10	FR/IDF/Paris	FR		http://iw3c2.org					
WWW 1994	1994						http://iw3c2.org					

Spreadsheet Editing via CSV is possible

- First – export your data to CSV
- ... then edit with a spreadsheet editor
- ... then import again
- Extension: Data Transfer
 - Import over csv
 - Good for well-formed input
 - But: no control over imported data
 - Many error/broken pages on ill-formed input
 - Follow-up issues to repair those pages
 - Spam possible → no validation
 - No check of template parameters

```
Title,Event[Acronym],Event[Series],Event[Ordinal],Event[Start date]
SMWCon 2020,SMWCon 2020,SMWCon,17,2020-11-24
SMWCon 2019,SMWCon 2019,SMWCon,16,2019-09-25
SMWCon 2018,SMWCon 2018,SMWCon,15,2018
SMWCon 2017,SMWCon 2017,SMWCon,14,2017
```

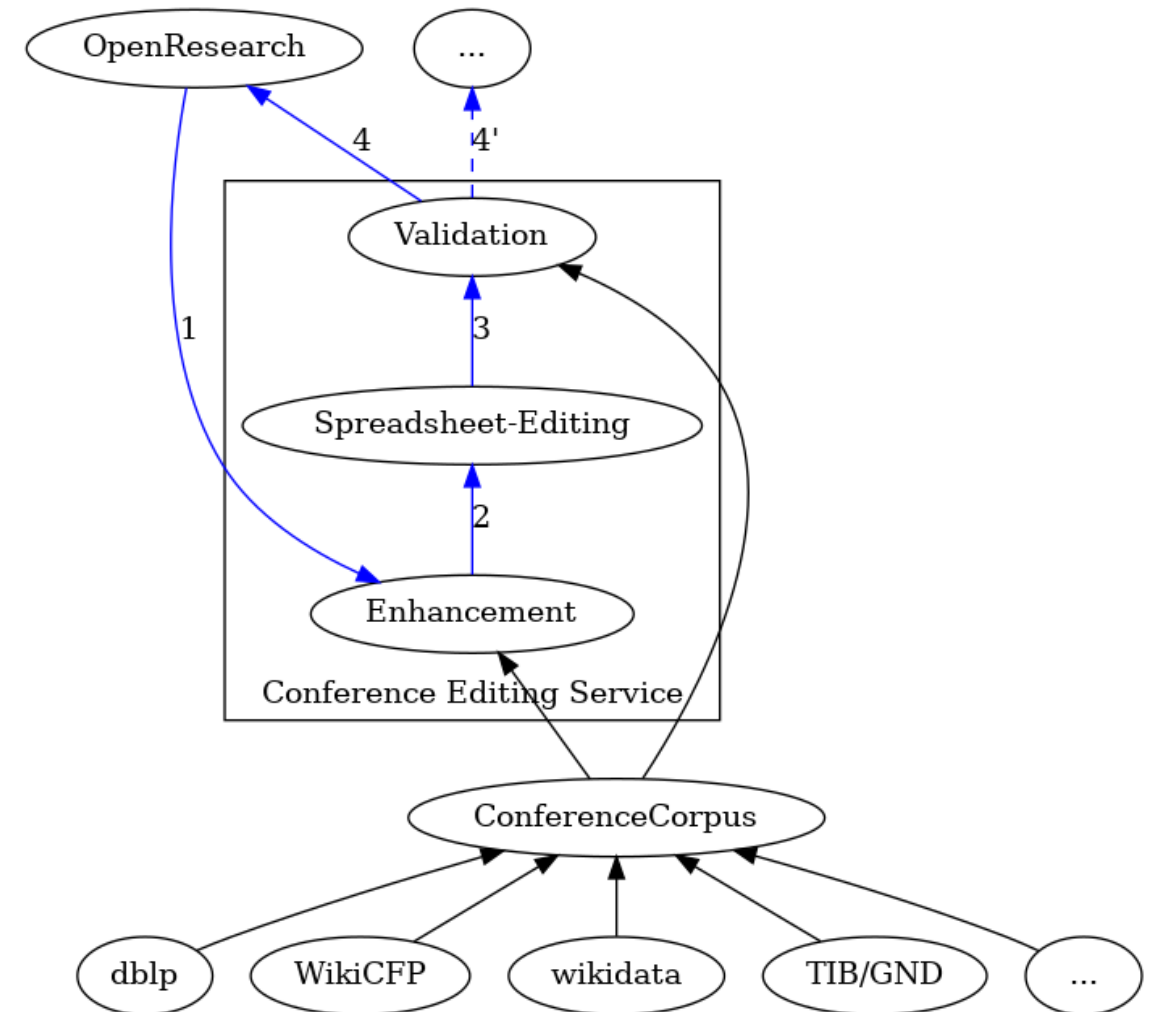
Improved approach: spreadsheet editing

- Download entities as spreadsheet
- Edit in favorite spreadsheet editor
- Upload changes
 - Validation and normalization is done automatically

The Web Conference 2020							
	A	B	C	D	E	F	G
1	pageTitle	Acronym	End date	Series	presence	Type	Country
2	WWW 1994	WWW 1995	1994-05-27	WWW		Conference	CH
3	WWW Fall 1994	WWW Fall 1995	1994-10-20	WWW		Conference	US
4	WWW 1995	WWW 1996	1995-04-14	WWW		Conference	DE
5	WWW Fall 1995	WWW Fall 1996	1995-12-14	WWW		Conference	US
6	WWW 1996	WWW 1996	1996-05-10	WWW		Conference	FR
7	WWW 1997	WWW 1997	1997-04-11	WWW		Conference	US
8	WWW 1998	WWW 1998	1998-04-18	WWW		Conference	AU
9	WWW 1999	WWW 1999	1999-05-14	WWW		Conference	CA
10	WWW 2000	WWW 2000	2000-05-19	WWW		Conference	NL
11	WWW 2001	WWW 2001	2001-05-05	WWW		Conference	CN
12	WWW 2002	WWW 2002	2002-05-11	WWW		Conference	US
13	WWW 2003	WWW 2003	2003-05-22	WWW		Conference	HU
14	WWW 2004	WWW 2004	2004-05-22	WWW		Conference	US
15	WWW 2005	WWW 2005	2005-05-14	WWW		Conference	JP
16	WWW 2006	WWW 2006	2006-05-26	WWW		Conference	GB
17	WWW 2007	WWW 2007	2007-05-12	WWW		Conference	CA
18	WWW 2008	WWW 2008	2008-04-25	WWW		Conference	CN
19	WWW 2009	WWW 2009	2009-04-24	WWW		Conference	ES

OpenResearch Use Case

- Incomplete series entries
 - 1:n relation between series and event
 - Missing or incorrect data
- Need to edit a complete series in one rush
- Validation on upload
 - Reduction of errors
 - Normalization of names
 - e.g. location names ("US", "USA", ...)



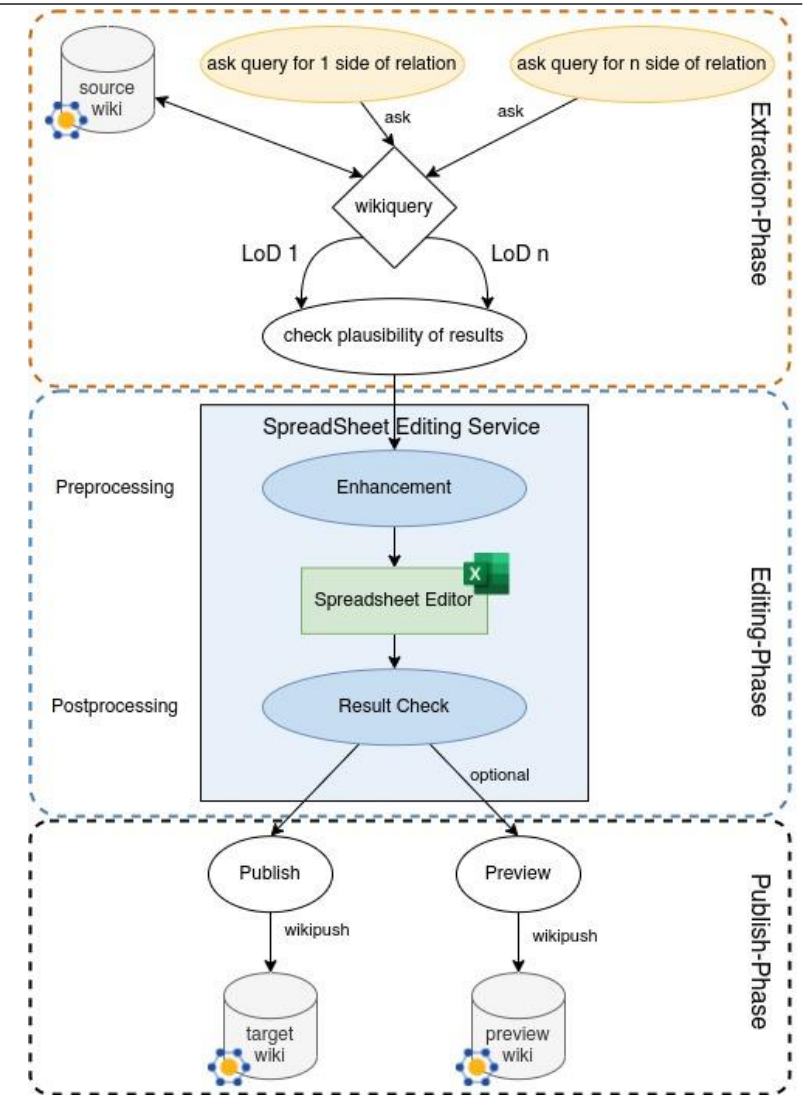
WWW example

- Edit and upload The Web Conference series

WWW	
The Web Conference	
Categories: World wide web	
http://www.iw3c2.org/conferences/	
wikidataId:	Q3570023
Scholia:	https://scholia.toolforge.org/event-series/Q3570023
DblpSeries:	www
WikiCFP Series:	3182
Bibliography:	dblp.uni-trier.de/db/conf/www/
CORE Rank (2017):	A*
CORE Rank (2018):	A*
Avg.	

What about ...

- Extraction-Phase:
 - Query wiki for data of the entities
 - I.e. extract properties from wiki-markup files
- Editing-Phase:
 - Convert entities to spreadsheet
 - Optionally:
 - Enhance data
 - Validate user input
- Publishing-Phase:
 - Apply changes to wiki-markup file
 - Upload changes



Limitations and Problems

- Properties set in template
 - Only queryable since not defined on page
- Properties defined over in-text annotation
- Access to properties of properties only possible with proper metadata management such as Pierre Racine's
- Open issues
 - Deletion of pages
 - Deletion of property values

GitHub



- Py-3rdparty-mediawiki
 - pywikibot
 - pycryptodome
 - mwclient
- Wikirender
 - wikitextparser
 - jinja2
- PyLoDStorage
 - sparqlwrapper
 - tabulate
 - pandas

Discussion

Feedback

What value would an OpenSource SMW Multipage Editing solution have four?
Please vote in the hopin platform ...

Realization

- Query page titles of the entities
 - I.e ask query
- Extract WikiSON
- Convert to spreadsheet
- Update WikiSON
- Upload changes

